

5

Filtro de Kalman Aplicado ao Modelo de Schwartz e Smith (2000)

A primeira parte deste capítulo, referente à passagem dos modelos estocásticos para as equações do Filtro de Kalman, já foi previamente abordada no Capítulo 2, mais especificamente na seção 2.2, na qual foi apresentado um resumo do artigo de SCHWARZ E SMITH (2000). Nesse sentido, os pontos abordados aqui são referentes a pontos que devem ser mais claramente expostos. A seguir é descrita a forma com que as variáveis de estado são estimadas. Por fim o método de maximização da função de verossimilhança através da otimização dos parâmetros do modelo proposto é explicitado.

5.1.

Passagem dos Processos Estocásticos para o Filtro de Kalman

O modelo de SCHWARTZ E SMITH (2000) pode ser descrito como um modelo de variações de curto-prazo e tendência de longo-prazo. As variações de curto-prazo são modeladas como um processo de reversão à média, revertendo para zero, e a tendência de longo-prazo como um movimento browniano. Além disso, os incrementos aleatórios são correlacionados.

$$d\chi_t = -k\chi_t dt + \sigma_\chi dz_\chi \quad (7)$$

$$d\xi_t = \mu_\xi dt + \sigma_\xi dz_\xi \quad (8)$$

$$dz_\chi dz_\xi = \rho_{\chi\xi} dt \quad (9)$$

O logaritmo do preço à vista do petróleo é dado pela soma das duas variáveis acima.

$$\ln(S_t) = \chi_t + \xi_t \quad (6)$$

Os processos estocásticos descritos nas equações (7), (8) e (9) apresentam valores esperados e matriz de variâncias-covariâncias dados por:

$$E[(\chi_t, \xi_t)] = [e^{-kt}\chi_0, \xi_0 + \mu_\xi t] \quad (10)$$

$$Cov[(\chi_t, \xi_t)] = \begin{bmatrix} (1 - e^{-2kt}) \frac{\sigma_\chi^2}{2k} & (1 - e^{-kt}) \frac{\rho_{\chi\xi} \sigma_\chi \sigma_\xi}{k} \\ (1 - e^{-kt}) \frac{\rho_{\chi\xi} \sigma_\chi \sigma_\xi}{k} & \sigma_\xi^2 t \end{bmatrix} \quad (11)$$

Portanto, o valor esperado e a variância do logaritmo do preço à vista podem ser escritos como

$$E[\ln(S_t)] = e^{-kt} \chi_0 + \xi_0 + \mu_\xi t \quad (12)$$

$$Var[\ln(S_t)] = (1 - e^{-2kt}) \frac{\sigma_\chi^2}{2k} + \sigma_\xi^2 t + 2(1 - e^{-kt}) \frac{\rho_{\chi\xi} \sigma_\chi \sigma_\xi}{k} \quad (13)$$

Como o objetivo é encontrar o valor esperado do preço à vista e não de seu logaritmo, é possível escrevê-lo como

$$\ln(E[S_t]) = E[\ln(S_t)] + \frac{1}{2} Var[\ln(S_t)]$$

Consequentemente

$$\begin{aligned} \ln(E[S_t]) &= e^{-kt} \chi_0 + \xi_0 + \mu_\xi t + \frac{1}{2} (1 - e^{-2kt}) \frac{\sigma_\chi^2}{2k} + \\ &\quad \sigma_\xi^2 t + 2(1 - e^{-kt}) \frac{\rho_{\chi\xi} \sigma_\chi \sigma_\xi}{k} \end{aligned} \quad (14)$$

5.1.1.

Utilizando a Medida de Martingale Equivalente Para Determinar a Relação Entre o Preço à Vista e Futuro

A ligação entre o preço à vista e o contrato futuro é feita através de processos estocásticos desenvolvidos sob medidas de probabilidade neutras ao risco. O preço do contrato futuro para certa data deve ser igual ao valor esperado do preço à vista para aquela data, logo

$$\ln(F_{T,0}) = \ln(E^*[S_T]) \quad (22)$$

E, portanto

$$\ln(F_{T,0}) = E^*[\ln(S_T)] + \frac{1}{2} Var^*[\ln(S_T)]$$

Utilizando os processos estocásticos convertidos para as novas medidas de probabilidade, ou seja, alterando suas tendências

$$\ln(F_{T,0}) = e^{-kT} \chi_0 + \xi_0 + A(T) \quad (23)$$

Onde:

$$\begin{aligned} A(T) &= \mu_\xi^* T - (1 - e^{-kT}) \frac{\lambda_\chi}{k} + \frac{1}{2} ((1 - e^{-2kT}) \frac{\sigma_\chi^2}{2k} + \\ &\quad \sigma_\xi^2 T + 2(1 - e^{-kT}) \frac{\rho_{\chi\xi} \sigma_\chi \sigma_\xi}{k}) \end{aligned}$$

5.1.2.

Discretização do Modelo Para Utilizar O Filtro de Kalman

As equações (7), (8) e (9) implicam que a equação de transição, ou equação de estado seja dada por

$$\begin{bmatrix} \chi_t \\ \xi_t \end{bmatrix} = \begin{bmatrix} 0 \\ \mu_\xi \Delta t \end{bmatrix} + \begin{bmatrix} e^{-k\Delta t} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \chi_{t-1} \\ \xi_{t-1} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} \quad (25)$$

Onde:

$$\text{Var}[\mathbf{v}_t] = \begin{bmatrix} (1 - e^{-2kt}) \frac{\sigma_\chi^2}{2k} & (1 - e^{-kt}) \frac{\rho_{\chi\xi} \sigma_\chi \sigma_\xi}{k} \\ (1 - e^{-kt}) \frac{\rho_{\chi\xi} \sigma_\chi \sigma_\xi}{k} & \sigma_\xi^2 t \end{bmatrix} \quad (26)$$

A equação (25) também pode ser escrita por duas equações distintas

$$\begin{aligned} \chi_t &= e^{-k\Delta t} \chi_{t-1} + v_{1t} \\ \xi_t &= \mu_\xi \Delta t + \xi_{t-1} + v_{2t} \end{aligned}$$

As equações acima, quando comparadas com as equações (10) e (11), permitem que se constate que são muito semelhantes, com a diferença do período inicial não ser mais o zero e sim o $t - 1$. O que as equações acima apresentam é na verdade, nos dois primeiros termos do lado direito de cada uma, o valor esperado para a variável em t , dado o valor da mesma em $t - 1$. Os últimos termos das equações inserem incerteza no processo. Os distúrbios, em cada uma, possuem a mesma variância de cada um dos processos e possuem a covariância também dada pela covariância entre os processos.

Desta forma, o processo inicia-se com valores estimados para as variáveis de estado e de sua matriz de variâncias-covariâncias dos erros de previsão. Posteriormente, são calculados os valores estimados por cada uma das equações que formam a equação matricial de transição. É importante ter em mente que as equações se ajustarão tão melhor às séries reais, quanto mais fidedigna for a estimativa dos parâmetros que regulam os processos estocásticos.

O outro grupo de equações, referentes à equação (23), pode ser escrito na forma do Filtro de Kalman como

$$\begin{bmatrix} \ln F_{T_1} \\ \vdots \\ \ln F_{T_n} \end{bmatrix} = \begin{bmatrix} A(T_1) \\ \vdots \\ A(T_n) \end{bmatrix} + \begin{bmatrix} e^{-kT_1} & 1 \\ \vdots & \vdots \\ e^{-kT_n} & 1 \end{bmatrix} \begin{bmatrix} \chi_t \\ \xi_t \end{bmatrix} + \begin{bmatrix} u_{1t} \\ \vdots \\ u_{nt} \end{bmatrix} \quad (28)$$

Ou ainda,

$$\ln F_{T_1} = A(T_1) + e^{-kT_1} \chi_t + \xi_t + u_{1t}$$

Até

$$\ln F_{T_n} = A(T_n) + e^{-kT_n} \chi_t + \xi_t + u_{nt}$$

As equações acima permitem que ao observar os contratos futuros no mercado seja possível estimar as variáveis de estado. No entanto, para que se obtenham as séries das variáveis de estado é necessário que se determine previamente estimativas para os hiperparâmetros do modelo. Os hiperparâmetros são aqueles desconhecidos no termo $A(T)$.

Através das séries de cotações dos contratos futuros, das estimativas dos hiperparâmetros, pelo critério de maximização da função de verossimilhança, que será abordado em seção próxima, e das estimativas iniciais para o vetor de variáveis de estado e para sua matriz de variâncias-covariâncias dos erros de previsão, o Filtro de Kalman pode ser rodado. O que significa que serão geradas séries para as variáveis de estado e para os erros das equações de estado e de medida. A aderência das séries geradas pelo Filtro pode, então, ser auferida através de medidas estatísticas que medem os erro de previsão em termos absolutos e relativos. Através do critério de convergência selecionado opta-se por estimar novamente os hiperparâmetros e rodar o Filtro, ou parar, pois a solução já é satisfatória.

5.2.

Aplicando a Estimativa de Máxima Verossimilhança ao Filtro de Kalman

A estimativa feita pelo critério de maximização da verossimilhança determina os parâmetros dos modelos para os quais a série de observações possui probabilidade de ocorrência máxima. O método parte do princípio de que uma amostra aleatória (y_1, y_2, \dots, y_T) de uma população tem uma função densidade de probabilidade (f.d.p.) dada por $f(\mathbf{y}, \boldsymbol{\theta})$, onde \mathbf{y} representa os valores da série e $\boldsymbol{\theta}$ os parâmetros que determinam o comportamento da série.

Portanto, para cada período de tempo em que se obtém uma observação existe também uma função densidade de probabilidade. O objetivo do método é, então, determinar para qual conjunto de parâmetros $\boldsymbol{\theta}$ o valor da f.d.p. conjunta, para todas as observações dos valores da série para cada tempo, se torna máximo. A f.d.p. conjunta será dada pelo produto das f.d.p individuais, para cada valor da série em um dado tempo, pois se sabe que a probabilidade de ocorrência de todos

os valores de uma dada série é o produto das probabilidades de ocorrência de cada um dos valores da série.

$$f(\mathbf{y}, \boldsymbol{\theta}) = \prod_{t=1}^T f(y_t, \boldsymbol{\theta}) \quad (51)$$

Onde:

$f(\mathbf{y}, \boldsymbol{\theta})$ representa a f.d.p. conjunta de ocorrência de todos os valores da série \mathbf{y} para os valores dos parâmetros $\boldsymbol{\theta}$.

$f(y_t, \boldsymbol{\theta})$ representa a f.d.p. de ocorrência do valor y_t dados os parâmetros $\boldsymbol{\theta}$ como geradores desta f.d.p.

Escrevendo de outra forma,

$$\prod_{t=1}^T f(y_t, \boldsymbol{\theta}) = f(y_1, \boldsymbol{\theta}) \times f(y_2, \boldsymbol{\theta}) \times \dots \times f(y_T, \boldsymbol{\theta})$$

A função acima é denominada função de verossimilhança. O objetivo econométrico passa a ser obter o vetor de valores estimados para os parâmetros da série de forma que a função de verossimilhança seja máxima.

$$L(\hat{\boldsymbol{\theta}}, \mathbf{y}) > L(\hat{\boldsymbol{\theta}}', \mathbf{y})$$

Onde:

L representa a função de verossimilhança

Ou seja, deseja-se obter o vetor $\hat{\boldsymbol{\theta}}$ para o qual nenhum outro vetor $\hat{\boldsymbol{\theta}}'$ apresente função de verossimilhança maior.

5.2.1.

Maximização da Função de Verossimilhança de Observações com f.d.p. Normal

O processo de maximização da função de verossimilhança é realizado através da derivação da função de verossimilhança em função dos parâmetros que compõem o vetor $\boldsymbol{\theta}$. O próximo passo é achar o vetor $\hat{\boldsymbol{\theta}}$ que resolve o sistema de equações formado pelas derivadas da função de verossimilhança igualadas a zero. Muitas vezes o processo de maximização da função de verossimilhança é feito em função do logaritmo neperiano da função de verossimilhança, pois o resultado geralmente é o mesmo, se a função for monotonicamente crescente, e a maximização se torna mais simples.

Se for considerada uma variável y , tal que, sua f.d.p é dada por:

$$y \sim N(\mu, \sigma^2)$$

Então

$$\hat{\mu} = \frac{\sum_{t=1}^T y_t}{T} \quad (52)$$

$$\widehat{\sigma^2} = \frac{\sum_{t=1}^T (y_t - \mu)^2}{T} \quad (53)$$

5.2.2.

Maximização da Função de Verossimilhança no Caso do Filtro de Kalman

Na seção anterior foi feita acima uma descrição do modelo de máxima verossimilhança para o caso em que as distribuições de probabilidades das observações da série apresentam distribuições normais e são independentes. No caso do Filtro de Kalman pode haver interdependência entre as observações, e é o que geralmente ocorre. Quando as observações não são independentes a função de distribuição de probabilidade conjunta se torna

$$L(\mathbf{y}; \boldsymbol{\theta}) = \prod_{t=1}^T p(y_t | \mathbf{Y}_{t-1})$$

Onde:

\mathbf{y} representa uma realização da série de variáveis observáveis

$\boldsymbol{\theta}$ representa os hiperparâmetros a serem estimados

$p(y_t | \mathbf{Y}_{t-1})$ é a f.d.p. para y_t dado que são conhecidas as observações da série até $t - 1$ e que a f.d.p. depende da série observada

Nos casos em que os ruídos da equação de medida (27) e da equação de estado (24), além da distribuição inicial para o vetor de variáveis de estado, dada por \mathbf{x}_0 e \mathbf{P}_0 , são normalmente distribuídos, a distribuição de y_t condicional a \mathbf{Y}_{t-1} também será Gaussiana. O vetor de variáveis de estado estimado também apresentará distribuição normal, dada por

$$p(\mathbf{x}_t | \mathbf{Y}_{t-1}) \sim N(\hat{\mathbf{x}}_t, \mathbf{P}_t)$$

É possível reescrever a equação (27) somando e ao mesmo tempo subtraindo $\mathbf{H}_t \mathbf{x}_{t-}$, da seguinte maneira

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{x}_{t-} - \mathbf{H}_t \mathbf{x}_{t-} + \mathbf{H}_t \mathbf{x}_t + \mathbf{u}_t$$

Colocando \mathbf{H}_t em evidência

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{x}_{t-} + \mathbf{H}_t (\mathbf{x}_t - \mathbf{x}_{t-}) + \mathbf{u}_t$$

O valor esperado de \mathbf{z}_t condicional a \mathbf{Z}_{t-1} será dado por

$$E(\mathbf{z}_t | t - 1) = \mathbf{H}_t \mathbf{x}_{t-} \quad (54)$$

E a matriz de variâncias-covariâncias será dada por

$$Var(\mathbf{z}_t | t - 1) = \mathbf{F}_t$$

Mas, pela definição de variância é possível escrevê-la da seguinte forma

$$\mathbf{F}_t = E \left[(\mathbf{z}_t - E(\mathbf{z}_t|t-1))(\mathbf{z}_t - E(\mathbf{z}_t|t-1))' \right]$$

Substituindo na equação acima o valor esperado de \mathbf{z}_t dado $t-1$ pela relação encontrada na equação (54), e \mathbf{z}_t pela equação de medida

$$\mathbf{F}_t = E [(\mathbf{H}_t \mathbf{x}_t + \mathbf{u}_t - \mathbf{H}_t \mathbf{x}_{t-}) (\mathbf{H}_t \mathbf{x}_t + \mathbf{u}_t - \mathbf{H}_t \mathbf{x}_{t-})']$$

Colocando \mathbf{H}_t em evidência nos dois termos do produto

$$\mathbf{F}_t = E [(\mathbf{H}_t (\mathbf{x}_t - \mathbf{x}_{t-}) + \mathbf{u}_t) (\mathbf{H}_t (\mathbf{x}_t - \mathbf{x}_{t-}) + \mathbf{u}_t)']$$

Obtendo a transposta do segundo termo

$$\mathbf{F}_t = E [(\mathbf{H}_t (\mathbf{x}_t - \mathbf{x}_{t-}) + \mathbf{u}_t) ((\mathbf{x}_t - \mathbf{x}_{t-})' \mathbf{H}_t' + \mathbf{u}_t')]$$

Aplicando a multiplicação distributiva dos termos

$$\mathbf{F}_t = E [\mathbf{H}_t (\mathbf{x}_t - \mathbf{x}_{t-}) (\mathbf{x}_t - \mathbf{x}_{t-})' \mathbf{H}_t' + \mathbf{H}_t (\mathbf{x}_t - \mathbf{x}_{t-}) \mathbf{u}_t' + \mathbf{u}_t (\mathbf{x}_t - \mathbf{x}_{t-})' \mathbf{H}_t' + \mathbf{u}_t \mathbf{u}_t']$$

Os valores esperados dos termos que possuem apenas um ruído serão iguais a zero, pois os valores esperados destes são zero

$$\mathbf{F}_t = \mathbf{H}_t E [(\mathbf{x}_t - \mathbf{x}_{t-}) (\mathbf{x}_t - \mathbf{x}_{t-})'] \mathbf{H}_t' + E [\mathbf{u}_t \mathbf{u}_t']$$

A variância dos erros de previsão das variáveis de estado é definida como

$$\mathbf{P}_{t-} = E [(\mathbf{x}_t - \mathbf{x}_{t-}) (\mathbf{x}_t - \mathbf{x}_{t-})'] \quad (55)$$

E a matriz de variâncias-covariâncias dos erros da equação de sinal é dada por

$$\mathbf{R}_t = E [\mathbf{v}_t \mathbf{v}_t'] \quad (56)$$

Substituindo as equações (56) e (55) na equação da matriz de variâncias-covariâncias

$$\mathbf{F}_t = \mathbf{H}_t \mathbf{P}_{t-} \mathbf{H}_t' + \mathbf{R}_t \quad (57)$$

Através das equações (54) e (57) é possível reescrever a equação da função de verossimilhança que foi determinada no Apêndice G como

$$\ln[L(\mathbf{y}_t; \mu, \sigma^2)] = -\frac{T}{2} \ln[2\pi] - \frac{T}{2} \ln[\sigma^2] - \frac{1}{2\sigma^2} \sum_{t=1}^T (\mathbf{y}_t - \mu)^2$$

O termo $\ln[L(\mathbf{y}_t; \mu, \sigma^2)]$ pode ser generalizado para $L(\mathbf{y}; \boldsymbol{\theta})$. O termo $-\frac{T}{2} \ln[2\pi]$ será exatamente igual, onde T representa o número de períodos para o qual se obtém observações. O segundo termo do lado direito da equação, dado por $-\frac{T}{2} \ln[\sigma^2]$, será alterado para $-\frac{1}{2} \sum_{t=1}^T |\mathbf{F}_t|$, pois a variância da série que era dada por σ^2 passou a ser fornecida pela matriz \mathbf{F}_t e poderá não ser constante. O último termo $-\frac{1}{2\sigma^2} \sum_{t=1}^T (\mathbf{y}_t - \mu)^2$ talvez seja o que sofre a alteração mais complexa,

passando a ser dado por $-\frac{1}{2}\sum_{t=1}^T \mathbf{u}'_t \mathbf{F}_t^{-1} \mathbf{u}_t$, onde $\mathbf{u}_t = \mathbf{z}_t - \mathbf{z}_{t|t-1}$, ou seja, exatamente igual ao termo usado anteriormente na equação acima. O mesmo motivo utilizado para o termo anterior, a não constância da matriz de variâncias-covariâncias, impede a saída dela do somatório, e como ela não é um escalar deve ser escrita na forma matricial conforme equação abaixo

$$L(\mathbf{y}; \boldsymbol{\theta}) = -\frac{T}{2} \ln[2\pi] - \frac{1}{2} \sum_{t=1}^T |\mathbf{F}_t| - \frac{1}{2} \sum_{t=1}^T \mathbf{u}'_t \mathbf{F}_t^{-1} \mathbf{u}_t \quad (58)$$

Desta forma, implementa-se um algoritmo que ao mesmo tempo em que roda o Filtro de Kalman maximiza a função de verossimilhança através da estimativa dos parâmetros do modelo. Uma forma de se montar o algoritmo é para um conjunto de parâmetros, definidos pelas derivadas parciais em função dos parâmetros ou por métodos de otimização numérica, no caso de problemas mais complexos, como é o caso do modelo de SCHWARTZ E SMITH (2000), nos quais as derivadas não possuem soluções fechadas, rodar o Filtro de Kalman e calcular a função de verossimilhança. O próximo passo é determinar novamente outro conjunto de parâmetros, determinados por algum método de otimização selecionado, e rodar novamente o Filtro e calcular a verossimilhança. O processo é repetido até que o ganho com a nova iteração passa a ser menor do que um padrão estabelecido, esta é a condição de parada.